

Standard Setting Approaches

Explanation and Considerations for Use

This tool will help schools familiarize themselves with the process for establishing a cutscore that decides whether a given graduation portfolio “passes” by demonstrating the desired level of proficiency.

This tool was created and/or compiled by The Rhode Island Department of Education and The Education Alliance at Brown University, with the generous support of the Bill & Melinda Gates Foundation.

<http://www.ride.ri.gov/highschoolreform/dslat/>
October, 2005



Standard Setting

In order to determine which Graduation Portfolios demonstrate student proficiency, schools and districts must undergo the process of setting performance standards. Setting performance standards is the process by which a performance standard or cut score is established¹. An assessment, such as a portfolio, can have a single cut score between two performance levels (e.g., “passing” and “not passing” or “proficient” and “not proficient”), or cut scores can be assigned between every performance level (e.g., Advanced, Proficient, Needs Improvement, and Failing). In essence, standard setting involves making a decision rule to classify students with adjacent scores differently (Cizek et al., 2004; Educational Testing Service, 2005). Standard-setting should not be confused with scoring, which is the process of assigning points to student work.

The following needs to be done before standard setting can take place:

1. Performance level categories, as well as performance level descriptors for each performance level, must be established and agreed upon.
2. Scoring of a large sample of student work (e.g, Graduation Portfolios) must occur *before* standard-setting can begin.

There are many standard setting methods in use today, and there is no agreed upon “gold standard” for standard setting. Nor does any performance assessment, such as a portfolio, have a “true” cut score. All standard setting procedures depend on the subjective judgment of a group of panelists who offer their recommended cut score to the individuals or entities that are legally authorized to set the “operational” cut score.

Some standard setting scores are more appropriate for the assignment of cut scores to bodies of student work that are performance-based, such as Graduation Portfolios. Regardless of the standard setting method chosen, however, all standard setting processes must include the following:

1. Select judges.
2. Teach judges about cutscores.
3. Define “borderline.”
4. Train judges in the use of the standard setting method.
5. Run cut score study.
6. Document results.
7. Set the legally operational cutscore.
8. Comply with the *Standards* on cutscores. (ETS, 2005)

These steps are described in more detail below.

¹ “Performance standards” are also commonly known as: achievement levels, classification scores, criterion levels, cut points, cut scores, cutscores, cutting scores, cutoff scores, mastery levels, passing scores, and performance levels (ETS, 2005).

Select judges.

Cutscore panelists, or judges, must be qualified to judge collections of student work put together by high school students. That is, they need to be knowledgeable both about the content areas assessed in a Graduation Portfolio and the population of students completing the Graduation Portfolio requirement. At the same time, they should be both representative of the student population and the community, as well as diverse. Finally, they must be acceptable to stakeholders in your district and willing to follow standard setting procedures.

Teach judges about cutscores.

Training of cut score judges is crucial. Before embarking on any standard setting, judges need to become thoroughly versed in the purpose of the Graduation Portfolio, the purpose of a cutscore, and the consequences of passing and failing the Graduation Portfolio requirement. To do so, they should learn about the student population in your district, the content of Graduation Portfolios, and the role of judges in the process of setting a cutscore. Finally, judges need to understand the two errors of classification:

- Classifying a student as proficient when s/he really is not.
- Classifying a student as non-proficient when s/he truly is proficient.

In fact, one of the aims of setting a cutscore is to “find the cutscore that causes the fewest misclassifications, or the least harm caused by errors of classification” (ETS, 2005).

Define “borderline.”

The next task in standard setting judge training is to define what “borderline” proficient portfolios look like. In other words, what kind of performance distinguishes between “basic,” “proficient” and “advanced” or, at the very least, between “proficient” and “non-proficient?” Assessment experts recommend that at least two hours of training be spent on defining borderline performance.

This activity must begin with a focus on the meaning of the performance levels and their descriptors. The starting points of such a discussion are the detailed descriptions in the district or state’s performance levels. Each sentence in the performance level descriptors should be analyzed and described. Next, to ensure that panelists attained a common interpretation of performance descriptors and the relationship of those descriptors to student work, panel members should individually assigned performance levels to a set of sample student Graduation Portfolios. Panelists then compare their individual results and discuss at length how the performance level descriptors supported their conclusions.

After a common understanding of the performance levels has been achieved, in-depth discussions of the performance of students who straddle the hypothetical line between

two performance levels (and their performance) needs to take place. A good way to start this dialogue is to review the description of the “proficient” performance level and then ask, “What does it mean to be just barely Proficient?” The intended result of such focused discussion is for all judges to have in their minds a picture of the same borderline student or portfolio before standard setting begins.

Train judges in the use of the standard setting method.

Training in a particular standard setting method should include these steps:

1. Overview of the standard setting method: What is it? Why do it? How is it done?
2. Practice using the standard setting method
3. Observe judges
4. Correct errors
5. Answer all questions
6. Practice more until all judges are prepared
7. Document judges’ understanding of the method. Have judges sign and date a statement that they have received training and understand the method. (ETS, 2005)

Run cut score study.

Methods used to set standards on collections of student work such as a Graduation Portfolio are called “holistic,” in that they require judges to focus judgment on a sample or collection of student work rather than a single test item or task at a time (Cizek, Bunch, and Koons, 2004). In holistic standard setting methods, 40 to 50 intact samples of student work such as Graduation Portfolios are selected to represent the range of student performance. These work samples have been scored prior to standard setting, but these scores are not initially shared with judges.

In the *Range-Finding* phase of standard-setting, identical sets of student work (i.e., Graduation Portfolios) are provided to each judge. Judges are asked to independently categorize the Graduation Portfolios based on the performance level descriptors, and without any discussion. This process reveals which Graduation Portfolios generate the most *agreement* and which generate the most *disagreement* between judges. The results are documented, and the Graduation Portfolios that generate the most *disagreement* define the score intervals in which the cutscores must fall. More Graduation Portfolios from those intervals are then selected for further evaluation in the next phase.

In order to determine final minimum scores for each performance level, judges next examine sets of work about which they disagreed in the range-finding phase, along with additional Graduation Portfolios representing those same score intervals. In this *Pinpointing* phase, judges assign performance levels to these Graduation Portfolios. The minimum score for each performance level is precisely “pinpointed” by determining the score around which there is, collectively, the maximum disagreement between judges.

This is the point that best represents the transition from Graduation Portfolios at a higher level to those at a lower level.

Document results.

From the very beginning of the standard setting process, it should be assumed that the cutscores that are set will need to be defended in court at some future point. Consequently, the entire cutscore training and setting process must be documented in language that you are willing to make public. All standard setting notes and other documents should be retained until legal counsel suggests that this is no longer necessary.

Set the legally operational cutscore.

Only a legally authorized entity (e.g., policy makers) can authorize the use a cutscore. Until this groups authorizes a cutscore that is derived from a standard setting procedure, it is known as the “study cutscore.” Once authorized, a cutscore becomes an “operational” cutscore.

Keep in mind, however, that the study cutscore does not always end up being the operational cutscore. The task of the legally authorized entity is to make the cutscore credible to the public; thus, the authorized group may change the study cutscore value to an operational cutscore that is more palatable to the public. Factors that may affect the value of the operational cutscore include:

- Relative harm caused by errors of classification
- Alignment with other assessments or course work
- Potential credibility to or acceptance by stakeholders
- Local policy issues
- Consequences or impacts on subgroups
- Overall purpose of the assessment

Comply with the Standards on cutscores.

The Standards of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education specify that the following be included in the standard setting process:

- Document how judges were selected, their qualifications and training, the procedures used, whether or not judges’ ratings were independent, the level of judge agreement, and how judges may have influenced one another (1.7)
- Score interpretations based on cut scores should be described and justified (4.4)
- The rationale and procedures for setting cutscores should be documented (4.19)
- When feasible, cutscores should be established on the basis of sound empirical data concerning the relation of test performance to relevant criteria (4.20)

Sources:

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Cizek, G.J., Bunch, M.B., Koons, H. (Winter 2004). Setting performance standards: Contemporary methods. Educational Measurement: Issues and Practice.

Educational Testing Service. (2005). Considerations in Setting Performance Standards (Cutscores). Montreal, Canada: National Council for Measurement in Education 2005 Training Session.

<http://www.aps1.net/District/MCAS/mcas2.htm>